Adaptive quarter-pel motion estimation and motion vector coding algorithm for the H.264/AVC standard

Seung-Won Jung, Chun-Su Park, Le Thanh Ha, and Sung-Jea Ko
Korea University, Department of Electrical Engineering, Anam-Dong, Sungbuk-Ku, Seoul, 136-701, Korea
E-mail: jungsw@dali.korea.ac.kr

Abstract. We present an adaptive quarter-pel (Qpel) motion estimation (ME) method for H.264/AVC. Instead of applying Qpel ME to all macroblocks (MBs), the proposed method selectively performs Qpel ME in an MB level. In order to reduce the bit rate, we also propose a motion vector (MV) encoding technique that adaptively selects a different variable length coding (VLC) table according to the accuracy of the MV. Experimental results show that the proposed method can achieve about 3% average bit rate reduction. © 2009 Society of Photo-Optical Instrumentation Engineers.
[DOI: 10.1117/1.3257262]

1 Introduction

In block-based video coders, motion estimation (ME) carries a great significance because of its impact on the compression efficiency. In order to achieve high compression efficiency, ME is performed with quarter-pel (Qpel) accuracy as well as half-pel (Hpel) and integer-pel accuracies in the H.264/AVC standard.[1] Even though ME with a high accuracy generally reduces the bits required for encoding the difference frame, it often compromises the total bit rate because the bits required for encoding the motion vector (MV) grows as the motion vector accuracy (MVA) increases.

Various methods of obtaining the optimal MVA has been introduced in the literature.[2,3] In Ref. 2, an optimal MVA is derived for each macroblock (MB) and for each frame. The optimal MVA formula in Ref. 2 reveals that the MVA is dependent on the texture and the interframe noise of the MB. In Ref. 3, the MVA is adaptively determined for each MB by examining all possible MVAs and selecting the one with the minimum Lagrange cost. However, the coding gain of these methods is limited, since additional bits indicating the MVA need to be encoded.

In this letter, a novel MVA decision algorithm for H.264/AVC is presented. The proposed method determines the validity of Qpel ME for each MB. Since no additional bit is required to indicate the MVA, the proposed algorithm can be implemented without modifying the syntax of the H.264/AVC standard. Then, in order to achieve the coding gain, we also propose an MV encoding technique that adaptively changes the variable length coding (VLC) table according to the MVA of the MB.

2 Proposed Algorithm

The proposed algorithm consists of two techniques. We first present an adaptive MV encoding technique for H.264/AVC. Then, based on the proposed MV coding technique, we also propose an MVA decision technique.

In H.264/AVC, not an original MV itself but the difference between the original MV and the predicted motion vector (PMV),[1] the motion vector difference (MVD), is encoded. Let ΔV denote the MVD defined as follows:

$$\Delta V = V_o - V_p, \qquad (1)$$

where V_0 and V_p represent the original MV and PMV, respectively. In H.264/AVC, each horizontal and vertical element of ΔV is independently encoded by using a common VLC table without considering the MVA of the MB.

For notational simplicity, we first define three motion vector sets:

$$MV_{all} = \{(u,v) | u,v \in 0, -1/4, 1/4; -1/2, 1/2, \dots \}, \qquad (2)$$

$$MV_{Qpel} = \{(u,v) | u,v \in 0, -1/4, 1/4, -3/4, 3/4, \dots \}, \qquad (3)$$

$$MV_{Hpel} = \{(u,v) | u,v \in 0, -1/2, 1/2, -1, 1, \dots \}. \qquad (4)$$

Note that MV_{All} is a union set of MV_{Qpel} and MV_{Hpel}. If ME is performed up to Hpel accuracy, the resulting MV should belong to MV_{Hpel}. If we allow Qpel ME, the additional MV set, MV_{Qpel}, is required to express the MV.

Assume that ME is performed up to Hpel accuracy for the current MB, i.e., $V_0 \in MV_{Hpel}$. Then, ΔV is an element of either MV_{Hpel} or MV_{Qpel} depending on V_p as follows:

$$\Delta V \in \begin{cases} MV_{Hpel}, & \text{if } V_p \in MV_{Hpel}, \\ MV_{Qpel}, & \text{if } V_p \in MV_{Qpel}. \end{cases} \qquad (5)$$

Note that the number of possible ΔV values is halved if Qpel ME is not applied. Therefore, instead of using a VLC table including all possible MVD values, VLC_{All}, a reduced-size VLC table containing either Hpel or Qpel MVD values, VLC_{Hpel} or VLC_{Qpel}, can be used. By adaptively changing the VLC table according to the MVA of the MB, the bits required to encode MVD in H.264/AVC can be effectively reduced.

The bits required for encoding the MV can be reduced by skipping Qpel ME. However, Qpel ME needs to be omitted when it has a negligible impact on ME performance. In order to determine whether Qpel ME is necessary, we consider two cases. In the first case (C1), we skip Qpel ME for all MBs and encode the MVDs using VLC_{Hpel}. We allow, in the second case (C2), Qpel ME for all MBs. Then, the resulting MVDs are encoded using VLC_{All}. By comparing C1 and C2, the effect of Qpel ME can be analyzed.

Let $dRDcost$ denote the difference between the rate distortion costs (RDcosts) of C1 and C2. If $dRDcost$ is positive, we can interpret that Qpel ME is required for the MB. This is because the loss caused by skipping Qpel ME is
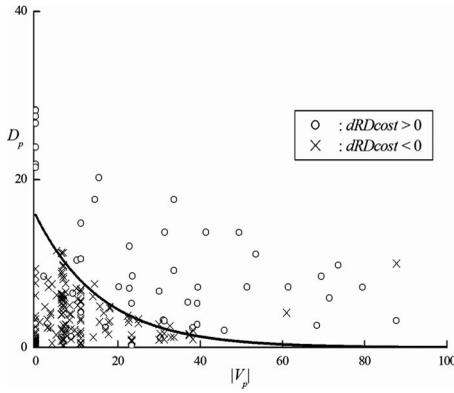
Fig. 1 Example of the *dRDcost* result for the Foreman sequence.

Table 1 Experimental conditions.

Software	JM 16.0 (Ref. 7)
Profile	High
Prediction structure	IPPP
Sequence resolution	CIF (352×288), 720p (1280×720)
Number of encoded frames	100
Number of reference frames	3
QP	32, 36, 40, 44
ME	Exhaustive (1/4 resolution)
ME search range	32 (CIF), 64 (720p)
Rate distortion optimization	Enabled
Entropy coding	CAVLC, CABAC

larger than the gain achieved by the proposed MVD coding technique. In the other case, Qpel ME can be considered unnecessary. Now, the remaining problem is to find the elements which affect *dRDcost*.

Motivated by the optimal MVA formula,[2] we claim that the necessity of Qpel ME increases as the spatial and temporal complexity of the MB increases. In our work, we estimate the spatial and temporal complexity of the current MB from the MB at the reference frame indicated by V_p, which is also available at the decoder. First, we measure the sum of the gradients of the luminance component as a spatial complexity metric.[4] Let ΔY_h and ΔY_v denote the horizontal and vertical gradient defined as

$$\Delta Y_h(x,y) = |\hat{Y}(x+1,y) - \hat{Y}(x,y)|,$$

$$\Delta Y_v(x,y) = |\hat{Y}(x,y-1) - \hat{Y}(x,y)|, \qquad (6)$$

where $\hat{Y}$ is the luminance component of the reference frame, and (x,y) is the pixel coordinate. Then, the spatial complexity of the MB, D_p, is obtained by averaging the gradient values inside of the MB as follows:

$$D_p = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \max[\Delta Y_h(x_p+i,y_p+j), \Delta Y_v(x_p+i,y_p+j)],$$

$$(7)$$

where N is the size of the MB, (x_p,y_p) is the pixel coordinate determined by V_p, and $\max(\cdot,\cdot)$ returns the maximum value between two values. The temporal complexity of the MB is simply defined as a magnitude of V_p,

$$|V_p| = (V_{p,h}^2 + V_{p,v}^2)^{1/2}, \qquad (8)$$

where $V_{p,h}$ and $V_{p,v}$ represent the horizontal and vertical elements of V_p, respectively.

Based on the preceding spatial and temporal complexity metrics, we examine the relation between *dRDcost* and complexity metrics. Figure 1 shows an example of the *dRDcost* result obtained by using the 80th frame of the Foreman test sequences. In this example, the quantization parameter (QP) is set to 36, one reference frame is used with CABAC coding, and the other experimental conditions are given in Table 1. For each MB, *dRDcost* is computed by performing C1 and C2, and *o* or *x* is assigned to

represent whether its value is positive or not. We can see that if the temporal or spatial complexity of the MB is high, Qpel ME tends to be advantageous. In the other case, where the spatial and temporal complexities are both low, the skipping of Qpel ME is beneficial in most cases. This tendency is consistent with the results in Ref. 2. To this end, we define an exponential curve to determine whether Qpel ME is advantageous:

$$f(|V_p|) = ae^{b|V_p|}, \qquad (9)$$

where a and b are modeling parameters. These parameters are obtained by taking logarithm function to Eq. (9) and applying the least-squares linear classification, so that a and b minimize the squared classification error.[5] By using test sequences in Ref. 6 and different QPs in Table 1, a and b are achieved by 9.29 and -2.99×10^{-2}, respectively. Then, Qpel ME is applied only when the estimated complexity point, $(|V_p|, D_p)$, is located above the curve of Eq. (9), where Qpel ME tends to improve the coding efficiency.

Figure 2 summarizes the procedure of the proposed algorithm at the encoder. When encoding the MB, D_p and
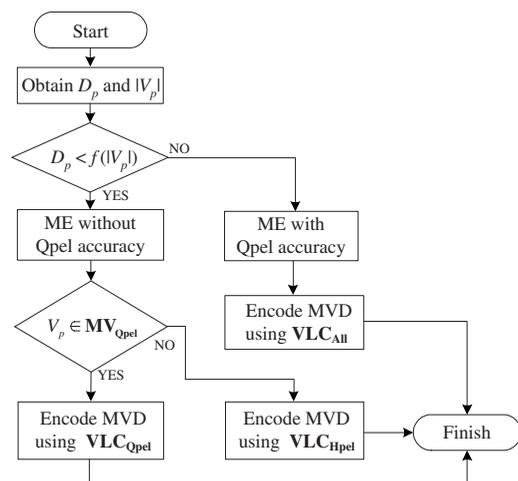

Fig. 2 Flowchart of the proposed algorithm.

Table 2 Performance of the proposed algorithm.

	Sequence	Ref. 2 (CABAC/CAVLC)			Proposed (CABAC/CAVLC)		
		$\Delta\text{bit}(\%)$	$\Delta T_E(\%)$	$\Delta T_D(\%)$	$\Delta\text{bit}(\%)$	$\Delta T_E(\%)$	$\Delta T_D(\%)$
CIF	Crew	−0.69/−0.28	0.75/1.24	0.15/0.04	−2.72/−2.03	−4.90/−4.54	−3.15/−4.15
	Foreman	−0.07/−0.19	1.01/0.19	0.28/0.21	−3.15/−3.15	−5.01/−4.17	−5.28/−4.58
	Salesman	−0.49/−0.18	−0.15/−0.36	0.17/0.09	−2.63/−2.18	−5.41/−4.45	−3.45/−1.67
	Soccer	−0.23/−0.42	0.23/0.22	0.13/−0.12	−2.65/−2.80	−3.91/−4.92	−4.93/−4.24
720p	Bigships	−1.12/−0.69	1.88/1.02	0.48/0.31	−2.11/−2.66	−1.97/−1.83	−5.78/−6.12
	City	−1.23/−1.36	0.39/1.03	0.75/0.41	−2.46/−2.08	−2.01/−1.89	−4.57/−3.99
	Jets	−1.20/−1.06	0.44/1.29	1.23/0.89	−3.62/−2.51	−2.31/−3.17	−4.78/−4.45
	Raven	−2.03/−2.70	0.92/0.35	0.69/0.98	−4.43/−4.72	−2.02/−2.32	−5.46/−4.96
	Average	−0.88/−0.86	0.68/0.71	0.48/0.34	−2.97/−2.77	−3.57/−3.41	−4.68/−4.27

$f(|V_p|)$ are calculated by using the PMV of INTER-16 ×16 mode, and the Qpel ME is applied if $D_p < f(|V_p|)$. This Qpel ME decision result is shared for all subpartitioned blocks. In other words, we do not allow each subpartitioned block to have different MVA. At the decoder, D_p is similarly obtained by Eq. (7). Then, if $D_p < f(|V_p|)$, the received MVD bits are decoded using either VLC_{Qpel} or VLC_{Hpel}, depending on the PMV. Since the same PMV should be used at both encoder and decoder, the PMV of INTER-16×16 mode is used at the decoder. In the other case, the MVD bits are decoded by using VLC_{All} as the conventional decoder.

3 Experimental Results and Conclusion

In order to evaluate the performance of the proposed algorithm, the proposed method is compared with the conventional algorithm in Ref. 3. A PC with an Intel Core2 Quad, 2.67-GHz CPU, and 8 GB RAM is used. The detailed experimental conditions are given in Table 1. The changes of Bjontegaard Delta (BD) rate (Δbit), encoding time (ΔT_E), and decoding time (ΔT_D) are used to measure the performance.[7]

Table 2 indicates that the proposed algorithm improves the coding efficiency of the original JM 16.0 by 2.97% and 2.77% for CABAC and CALVC, respectively.[8] Since the proposed algorithm does not require any overhead bit to indicate the MVA, superior coding efficiency is obtained when compared to the conventional algorithm. Here, it should be noted that the Bigships and Jets sequences are encoded by using 151th to 250th and 301th to 400th frames, where shot changes occur, respectively. Since the intracoding outperforms the intercoding in such cases, the performance of the proposed algorithm is not deteriorated.

From the viewpoint of the computational complexity, since all MVAs are examined and the best one is selected,[3] the complexity of the original JM 16.0 encoder and decoder is maintained or slightly increased. In the proposed algorithm, by skipping unnecessary Qpel ME, additional computation for D_p at the encoder is compensated and even a slight encoding time saving of 3.57% and 3.41% is achieved by CABAC and CAVLC, respectively. In addition, although the decoder should compute D_p for each MB, the decoding time is also saved by 4.68% and 4.27% on average for CABAC and CAVLC, respectively. This is because the interpolation time for the motion compensation is decreased due to the skipping of the unnecessary Qpel ME at the encoder.

In this letter, we first presented an adaptive MVD coding scheme. Then, in order to apply the adaptive MVD coding technique effectively, we also proposed an algorithm that selectively performs Qpel ME based on the spatial and temporal complexity of the MB. The experimental results demonstrated that the proposed algorithm improves coding efficiency without requiring the computational overhead at both encoder and decoder.

References

1. T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 560–576 (2003).
2. J. Ribas-Corbera and D. L. Neuhoff, "Optimizing motion vector accuracy in block-based video coding," *IEEE Trans. Circuits Syst. Video Technol.* **11**(4), 497–511 (2001).
3. J. Shen and J. Ribas-Corbera, "Benefits of adaptive motion accuracy in H.26L video coding," *Proc. IEEE ICIP* **1**, 1012–1015 (2000).
4. A. R. Varkonyi-Koczy, A. Rovid, and T. Hashimoto, "Gradient-based synthesized multiple exposure time color HDR image," *IEEE Trans. Instrum. Meas.* **57**(8), 1779–1785 (2008).
5. J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.* **9**, 293–300 (1999).
6. T. Tan, G. J. Sullivan, and T. Wedi, "Recommended simulation common conditions for coding efficiency experiments revision 4," ITU-T SG16/Q.6 Doc. VCEG-AJ10, San Diego, CA (2008).
7. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG16/Q.6 Doc. VCEG-M33, Austin, TX (2001).
8. K. Sühring, JM 16.0 reference software. Available at http://iphome.hhi.de/suehring/tml/download/.